

リレーショナル型標本データベースの開発 — 早稲田大学が所有する地質標本でのテストケース —

関口 寿史・円城寺 守

Exploitation for the Sample Database of the Relational Type
— as a Test Case for the Geologic Samples in Waseda University —

Hisashi SEKIGUCHI and Mamoru ENJOJI

要 旨

早稲田大学教育学部理学科地球科学専修が所有する地質標本を管理するための標本管理システムを開発した。このシステムは、従来に多くみられるカード型データベースよりも相互関連性を大幅に強化したリレーショナル型データベースを基幹とするシステムであり、主な操作はネットワークを介して Web 上で利用することができる。各種の鉱石標本・岩石標本・鉱物標本を相互関連性の強化によって体系的に管理し、文字や数値情報だけでなく画像やデータファイルなどを柔軟に収容できるようにした。

システム構成は、データの保管を担うバックエンドシステムとデータ処理を行うフロントエンドシステムとからなる。バックエンドシステムはデータベースサーバに MySQL、ファイルサーバに NFS 共有を、フロントエンドシステムは Web アプリケーションサーバとして Apache と PHP を使用している。以上のアプリケーションが稼働する OS には、CentOS Linux を使用した。これらのオープンソースソフトウェアを利用することで、開発と運用のコストを大幅に抑えることが可能となった。また、バックエンドとフロントエンドを分離したことで、管理用ページや検索ページ等、目的に応じた Web フォームを用意することができる。

今回のシステム開発により、標本管理システムにおいてリレーショナル型データベースが有用であることが確かめられ、研究・教育においてオープンソースソフトウェアが有効に利用できることが期待される。

キーワード：地質標本，標本データベース，オープンソース，RDBMS，Web

1. はじめに

早稲田大学教育学部理学科地球科学専修が所蔵する標本は、整理・登録されたものだけでも優に4,000点を超え、また整理途上にある研究試料も多く存在する。これらの中には太平洋戦争以前に採取されたものもあり、国土開発による整地や鉱山の閉山、戦後の海外情勢の変化などによって、現在では入手が困難あるいは不可能になってしまったものも多く含まれている。また、地質標本は物理化学における標本と異なり、地球上で唯一無二の存在である（柳沢ほか，1998）。有用有効な標本を適正に維持し管理することの重要性は今後も益々大きくなるであろう。これらの貴

重な標本を効率よく管理できる方法を模索した結果、この小論で紹介する標本データベースの開発に至った。

今日では、全国の研究・教育機関において様々な地質標本管理システムが稼働している。たとえば、地質調査所（現・独立行政法人 産業技術総合研究所）においては、GEMS が1970年代から開発され、運用され続けている（坂巻・小野，1976）。1990年代以降のインターネットの普及によって、Web 上で利用できる検索システムもみられるようになってきた（山本ほか，2002）。また、リレーショナルデータベース管理システム（Relational Database Management System; RDBMS）が市場に出回るようになってからは多くの標本管理システムが RDBMS 上に展

開されるようになった。しかしそれらは依然としてカード型データベースとして運用されており、必ずしもRDBMSを活用しているものではなかった (Fig. 1 a, b)。

今回開発した標本データベースの特徴は、テーブル同士の相互関連性を強化したデータ構造 (Fig. 1 c) と、データの入力・出力・検索等のすべての操作をネットワークを介した Web によって適切な権限で実行できることである。さらに以下の点を考慮した。

1) オープンソースソフトウェア (Open Source Software; OSS) を積極的に採用し、開発・運用コストを下げる。

2) バックエンドシステム (データベースシステム、ファイルシステム) とフロントエンドシステム (Web アプリケーションサーバ) とを独立させ、データ入力フォームや検索フォームなど、目的に特化した Web を用意できる (Fig. 2)。

a. 従来のカード型 (1)

標本テーブル

分類	標本名	産地	...
鉱石	金鉱石	茨城県
岩石	蛇紋岩	埼玉県
鉱物	石英	山梨県
・	・	・	・
・	・	・	・

b. 従来のカード型 (2)

(テーブルを分離)

鉱石標本テーブル

標本名	産地	...
金鉱石	茨城県
・	・	・
・	・	・

岩石標本テーブル

標本名	産地	...
蛇紋岩	埼玉県
・	・	・
・	・	・

鉱物標本テーブル

標本名	産地	...
石英	山梨県
・	・	・
・	・	・

c. リレーショナル型

標本テーブル

標本名	産地	...
金鉱石	茨城県
石英	山梨県
蛇紋岩	埼玉県
・	・	・
・	・	・

相互関連性

鉱床テーブル

鉱床名	所在地	...
栃原鉱床	茨城県
・	・	・
・	・	・

鉱物テーブル

鉱物名	化学組成	...
自然金	Au	...
石英	SiO ₂	...
・	・	・
・	・	・

岩石テーブル

岩石名	分類	...
蛇紋岩	変成岩	...
・	・	・
・	・	・

Fig. 1 データベース構造例

a. 従来のカード型データベースであり、すべての標本を単一のテーブルで扱い、個々のレコードはカードとして呼び出している。b. a と基本的には同様で、標本の分類によってテーブルを分割している。c. リレーショナル型データベースを活用し、包括するデータに適したテーブルを用意して、それぞれのレコードに対してリレーションシップを与えてある。情報の呼び出し方は様々である。

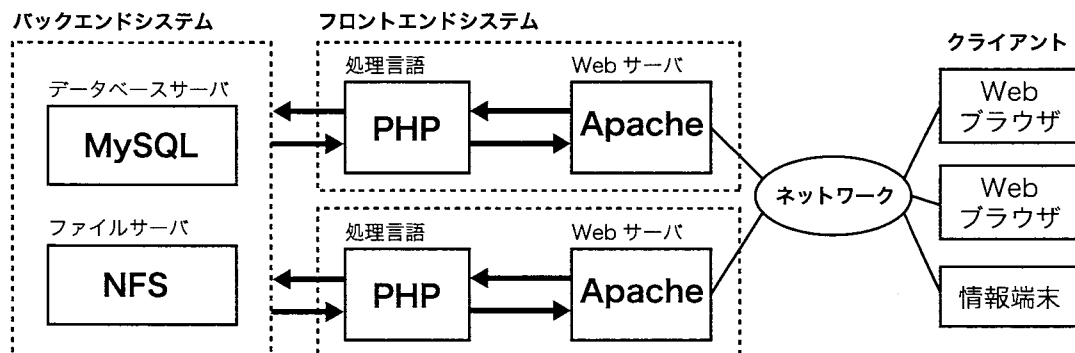


Fig. 2 システム構成図

データ保管を担うバックエンドシステムと、処理系のフロントエンドシステム（複数）からなるデータベースシステムである。クライアントはそれぞれの端末仕様や目的に応じたフロントエンドシステム（Web）にアクセスすることになる。

3) 文書・画像・ファイル等の様々な形態の情報の保持を可能にし、データの保存に柔軟に対応する。

この標本データベースは研究試料管理も含めた統合データベースの一部分の機能として稼働している。地球科学専修には、各種の地質標本や化石標本が収蔵されているが、この小論では、岩石・鉱石・鉱物の標本管理部分に限って述べる。このデータベースシステムのさらに詳しい仕様等については別途報告する予定である。

2. システム構成

標本データベースは、データを保持するためのバックエンドシステムとなるデータベースサーバおよびファイルサーバと、要求を処理するためのフロントエンドシステムとなる Web アプリケーションサーバとから構成される (Fig. 2)。負荷分散と冗長性確保のために役割に応じて複数台のサーバでの運用が好ましいが、1 台のサーバで運用することも十分可能である。現在運用しているハードウェアは、HP 製の x86 アーキテクチャのサーバ 1 台であり、VMware ESXi を使用した仮想環境の上で前述の複数のシステムを稼働させている。OS には CentOS Linux を使用し、サーバアプリケーションは、Web サーバに Apache、処理言語に PHP、データベースに MySQL、ファイルサーバには NFS 共有をそれぞれ使用している。以上のソフトウェアはすべて OSS で賄うことができ、ライセンス条項に準拠していれば無料で利用することができる (VMware ESXi は OSS ではないが無償で利用することができる)。

サービスにアクセスするためのクライアントは、サーバ側のフロントエンドシステムの仕様に依存するが、基本的

には一般的なブラウザ (Microsoft IE, Apple Safari, Google Chrome, Mozilla Firefox 等) とネットワーク接続を備えていれば問題はない。

3. テーブル設計

標本データベースのテーブルは、標本・鉱床・岩石・鉱物を管理する主テーブルと、主テーブルのデータを補うための副テーブルとからなり、それぞれのテーブルがリレーションを持ち、データベースシステムを形成している (Fig. 3)。

3.1. 主テーブル

主テーブルは、基本となる標本テーブルと鉱床テーブル・岩石テーブル・鉱物テーブルからなる。標本テーブルには鉱床テーブル・岩石テーブル・鉱物テーブルの各テーブルとのリレーションが任意に与えられ、一つの標本テーブルで鉱石標本・岩石標本・鉱物標本のいずれにも対応する (Fig. 1 c)。また上の構造から、標本データベース、鉱床データベース・岩石データベース・鉱物データベースの面をもち、それらが互いに連携している統合データベースとすることもできる。

3.1.1 標本テーブル

標本データベースの核となる標本テーブルには、標本種別・識別番号・採取地・採取者・配架 (収納) 場所等、標本が有する固有の情報が与えられる。本来、鉱床や岩石、鉱物の各情報は標本に固有しないため、鉱床テーブル・岩石テーブル・鉱物テーブルとして標本の個々のレコードとのリレーションを任意に保持する。採取位置情報は、緯度

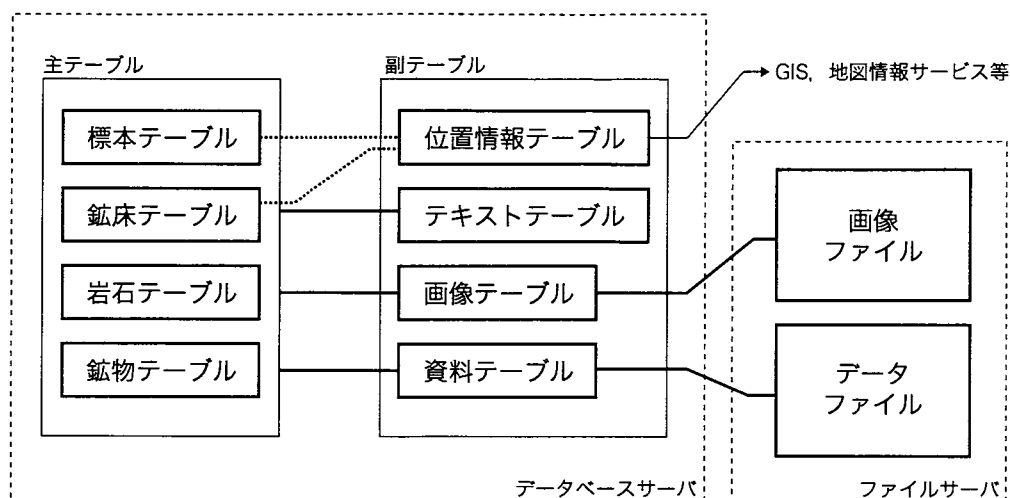


Fig. 3 データ構造

主テーブルに対して副テーブルのリレーションが与えられ、主テーブルの情報を補完する。それぞれの副テーブルはすべての主テーブルへのリレーションを持つことができるが、位置情報テーブルのリレーションは位置情報が必要な標本テーブルと鉱山テーブルのみになる。画像ファイルおよびデータファイルなどのファイル形式のデータはデータベーステーブルとともにファイルサーバへ保存される。

と経度による表記 (WGS-84座標系) および地名表記からなり、さらに後者は国名・都道府県名・市区町村名と採取当時の産地名をラベル表記してある。標本の収納位置としての配架場所には3つの階層を用意し、部屋番号・棚番号・引出番号といった利用ができる。

3.1.2 鉱床テーブル

鉱床の一般的な情報を保持する鉱床テーブルは、鉱床名、鉱床の型、産出鉱種、鉱石品位・鉱量、開山・閉山等の沿革、位置情報等を与える。位置情報は標本テーブルと同様の情報をもつ。

3.1.3 岩石テーブル

岩石テーブルは、当該岩石の一般的な情報を与えるほか、関連する文献情報・化学分析値・年代情報なども含んでいる。それぞれの岩石は、系統樹状の構造を呈し、火成岩 - 深成岩 - 花崗岩 - ...等の階層構造で示すことができる。

3.1.4 鉱物テーブル

鉱物テーブルは、鉱物の一般的な情報を含んでいる。それらは、鉱物名・化学組成・分類・結晶系・形態・色・屈折率・硬度・劈開等の性質である。それぞれの鉱物は岩石と同様に系統樹状の構造も呈し、長石 - アルカリ長石 - 正長石 - ...等の階層構造を持つことができる。

3.2 副テーブル

主テーブルに対してテーブル外の情報を保持する手段と

して副テーブルを用意してある。副テーブルには、テキストテーブル・画像テーブル・資料テーブルがあり追加情報として様々な形態のデータを主テーブルのレコードに対して幾つでも付加することができる (Fig. 3)。また、位置情報も内部では別テーブルとして管理しており、GIS や地図情報サービスとの親和性を考慮してある。

3.2.1 テキストテーブル

テキストテーブルは、形式化された主テーブルの情報を補完するために、タイトルと長文のテキストを用意している。例えば、当該標本に固有の産状や地質学的意義、地域的歴史的背景などを文章として書き込むことができる。また収蔵標本は卒業論文、修士論文、博士論文、公表論文などの素材であるものも多く、これらの情報も標本に対して与えられる。テキストは書式無しのプレーンテキストまたはHTML 書式付きで保存され、項目数を任意に増やすことができる。

3.2.2 画像テーブル

画像テーブルは、写真や図などの画像ファイルを管理するテーブルであり、題目やコメント・スケール・撮影情報等を有している。1枚の画像には階層として複数の画像ファイルが割り当てられ、切り替えたり重ねたりして表示できる。階層構造を採用することで、例えば光学条件を変えて撮影した同一標本の種々の顕微鏡画像を保存したり、スケールや説明を透明階層として重ねて原画像に変更を加えるこ

となく保存したりすることができる。画像の保存に際してはサムネイル表示用の縮小画像を内部で自動的に生成し、閲覧時の通信・処理負荷を低減させている。スケール情報は1インチあたりのピクセル数を数値として保存し、閲覧時にスケールバーを生成したり画幅の実寸値として表示したりすることができる。保持できる画像の数は任意であり、その中から一覧表示のサムネイルとして代表画像を選択する。画像ファイル自体はファイルサーバに保管される。

3.3.3 資料テーブル

資料テーブルは添付したデータファイルを管理するテーブルであり、題目やコメント、ファイル形式情報等を保持する。画像や音声など様々な形式のファイルを元の情報を残したまま保管できる。添付できるファイルサイズは、サーバの仕様にもよるが、1ファイルあたり2GB程度の容量には十分に対応でき、保持できるファイル数も任意である。データファイル自体はファイルサーバに保管される。

4. 運用

この標本データベースに収録されているデータは、従前より存在していた標本台帳を電子化した単一の一覧表データである。これらの表データを、言語処理あるいは手作業で、鉱床リスト・鉱物リスト・岩石リストに相互関連性を維持しつつ分離し、データベースに転送した。また、元となったデータには採取地等の表記に統一性がないものもあり、情報を失わない程度に最適化してある。

現在テスト運用中のシステムには、Web フロントエンドとして、データの入出力等を行う管理 Web ページと検索表示を行う検索 Web ページとを用意してある (Fig. 4)。

管理 Web ページはデータの追加や修正等の標本管理を行うためのページであり、ID とパスワードでのログインを前提にしている。各データ (標本・鉱床・岩石・鉱物) の各操作 (表示・追加・修正・削除) は各ユーザに独自に割り当てることができ、不用意なデータの改変を防ぐことができる。 (Fig. 4 a)

検索 Web ページは、Ajax を利用してリアルタイム表示を可能にした検索システムである。目的が検索のみであるため、データベースアクセスを読み取りに制限し、機能も管理 Web から大幅に削減して簡素化と高速化を図った。この Web では、検索文字を入力する度に、該当する項目をリアルタイムで表示できるようにしてあり、検索効率を向上させてある。 (Fig. 4 b)

情報検索では、鉱床・岩石・鉱物のテーブルを単一化したため、それらの種別間の標本をシームレスに検索できる。かつ、鉱床・岩石・鉱物テーブルの相互関連性強化によって、それぞれのテーブルから標本を検索したり、互いに関

連する情報について体系的に横断したりすることもできる (Fig. 5 a)。また、位置情報を別テーブルで管理したことで、地図上に採取位置や鉱山位置をポイントし、該当する標本を探すこともできる (Fig. 5 b)。

5. まとめと今後の展開

情報技術 (IT; Information Technology) の発達により、情報の器としてデータベースシステムは様々な分野で活躍している。さらに近年では、通信技術の向上や低廉化とも相まって、世間の潮流は情報通信技術 (ICT; Information and Communication Technology) に遷移してきており、情報技術と通信技術を融合活用した集合知の構築に重点を置くようになってきた。

数年前までの多くのサーバ OS やソフトウェアは大層高価であり、必然的に大手市場での利用に限られ、研究や教育目的で使用するには敷居が高かった。しかし近年では、Linux や MySQL データベースのような高性能なソフトウェアがオープンソースソフトウェア (OSS) として公開され、アイデア次第で自由に開発が行えるようになった。特に研究や教育目的における OSS の存在意義は大きい。また OSS は官公庁での採用例も増えてきており、今後も盤石な展開が期待できる。

以上の点から、RDBMS をはじめとする OSS とネットワークを有効に活用した標本データベースの開発は、今後の標本管理システムに対して大いに有効であることが期待できる。

現在のテスト運用にあたって現れた課題は、分類や語句等の統一と情報の共有に際しての標準化である。鉱床型や岩石型、地層名等には様々な分類や表記が存在し、研究者や学界の流儀によっても異なる場合がある。旧来よりこの問題は標本整理における課題として議論され続けてきた (柳沢ほか, 1998)。最近では地学用語の標準化の動きもあり、今後は利便性を見計らいながら随時これらを採用する予定である。さらに自然言語処理等の高度な言語処理が OSS として利用できる現在では、地学版のシソーラス辞書を用意し、あいまい検索して利用する方法も考えられる。また前述の集合知の構築のためにも、インターネットを活用し、全国各地に点在する標本データベースを横断検索できるシステムの構築が今後必要になってくるだろう。そのためにも、標本データベースでの通信仕様の標準化が望まれる。上述の課題の他にも、さらなるリレーションの強化など、データベース構造について追求すべき点は多い。

以上で紹介した標本データベースシステムは現在、早稲田大学の学内ネットワークにてテスト運用中である。今後、

a. 管理 Web ページ

管理 Web ページ

金鉱石 gold ore

DANA No. 307 金表示標本

標本識別記号 TH45016

化学組成 Au

採取地 JAPAN

採取者 丸山 久志郎 大子町 新屋

調査者 関口 守

調査地 新屋金山

関連岩石/鉱物 自然金

配製場所 自然金

写真 (編集)

金鉱石 gold ore

b. 検索 Web ページ

検索 Web ページ

検索結果: 標本247件, 鉱物10件, 岩石10件, 鉱山10件

標本	名称	種類	産地
1010101	gold	自然金	KOR 金井高山
1010102	gold	自然金	JPN 秋田県 鹿角郡
1010103	gold	自然金	JPN 鹿児島県 船橋町
1010104	gold	自然金	JPN 兵庫県
1010105	gold	自然金	JPN 埼玉県 大黒坑
1010106	gold	自然金	I 自然金
1010107	gold	自然金	KOR 全羅北道 月田里金山
1010108	gold	自然金	TWN 金瓜石鉱山
1010109	gold	自然金	JPN 兵庫県 赤穂郡船板村桶見坑
1010110	gold	自然金	I 自然金
1010111	gold	自然金	KOR 平安南道 大同郡龍平面上次里
1010112	gold	自然金	JPN 岩手県 紫波郡佐比内村
1010113	gold	自然金	I 自然金
1010114	gold	自然金	KOR 忠清北道 槐山郡沙梁面谷里二号脈
1010115	gold	自然金	北海西造
1010116	gold	自然金	JPN 鹿児島県 船橋町
1010117	gold	自然金	I 自然金

Fig. 4 Web ページの例 (1)

a. 標本を管理するための Web ページであり、データの追加・修正・削除等のすべての操作を行うことができる。画像データや資料などのファイルデータもこのページから追加できる。この管理ページにはログインが必要であり、許可されたユーザのみが操作可能である。b. 標本を検索するための Web ページであり、文字列、鉱床種等のカテゴリー、地図上での絞り込み検索ができる。また、このページは Ajax を利用しており、検索条件を反映しながらリアルタイムに表示することができる。

a. 関連情報（鋤床）

[illegible]

b. 地図上へのマッピング

新刊情報 森羅	管理ツール ホーム ログイン ログアウト
TOP	ログインユーザー名: admin ID: 12345 [Logout]
最新記事	地図 63件を表示しました [詳細表示]
地学一般	
火山 岩石 鉱物	分類 <input type="text" value="鉱山"/>
地球科学教室	ポイント <input type="button" value="ポイントを選択"/>
標本	関連情報
地球学研究室	投稿情報 <div style="background-color: #f0f0f0; padding: 10px;"> NO PHOTO <div> 掲載状況 公開済 / 未承認 / 削除済み / 待機中 </div> </div>
試料 調査	
月報	
文献 歴史 位置 地図	
管理ツール	
ホーム 写真 ファイル サイト管理 ユーザ管理	

Copyright © 2009 SK-MAC. Geospatial Consulting. All Rights Reserved.

Fig. 5 Web ページの例 (2)

a. 例えば鉱床のデータには関連情報として該当する標本が表示され、これらのリンクを辿ることで体系的に横断できる。標本・岩石・鉱物にも同様の関連情報が与えられている。
b. 地図上に位置情報をマッピングすることで、地図範囲から標本や鉱床を検索することができる。また、地図上のポイントををクリックすることで、その位置に関連した情報を辿ることができる。（図中の地図の表示には Google Maps API を利用している。）

計画中の博物標本展示室（仮称）のオープンに向けて公開予定である。さらなる可用性を向上させ、また継続的に情報を充実化させ、貴重な標本の有効利用に供したい。

引用文献

- 松江 千佐世・遠藤 祐二・兼子 尚知・青木 正博（2004）：地質標本登録データベース「日本産鉱石データ」の公開 - 産総研研究情報公開データベース（RIO-DB）課題として -, 情報地質, 15, (3), 164-167.
- 柳沢 幸夫・松江 千佐世・牧本 博（1998）：地質標本データベース研究の現状. 地質ニュース, no. 532, 41-48.
- 坂巻 幸雄・小野 晃司（1976）：あたらしい標本管理・検索システム〜GEMS〜の誕生. 地質ニュース, no. 265, 30-35.
- 山本 嘉一郎・川村善也・西脇 二一・神谷 英利（2002）：在日本脊椎動物化石標本データベース JAFOV の Web サービス. 情報地質, 13 (3), 175-178.